




INCASI Working Paper Series

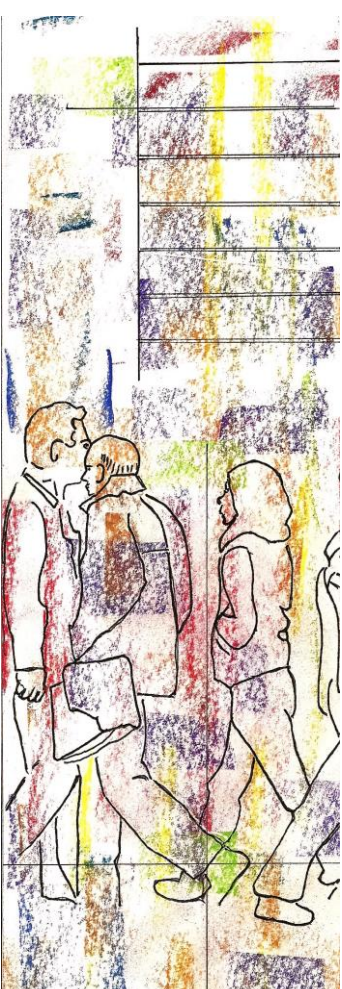
2018, No. 2

 **INCASI** *International Network for
Comparative Analysis of Social Inequalities*



Introducción a la lógica multinivel: un análisis longitudinal con SPSS y R

Mattia Vacchiano
Jordi Merino Noé



European
Commission

Horizon 2020
European Union funding
for Research & Innovation

Marie Skłodowska-Curie Actions (MSCA)
Research and Innovation Staff Exchange (RISE)
H2020-MSCA-RISE-2015
GA-691004

Introducción a la lógica multinivel: un análisis longitudinal con SPSS y R

Mattia Vacchiano¹
Jordi Merino Noé²

¹ Centre d'Estudis Sociològics sobre la Vida Quotidiana i el Treball (QUIT)
Institut del Treball (IET)
Universitat Autònoma de Barcelona, Spain
mattia.vacchiano@uab.cat

² Unitat de Sociologia. Departament d'Empresa.
Universitat de Girona, Spain
jordi.merino@udg.edu

INCASI Working Paper Series is an online publication under *Creative Commons* license. Any person is free to copy, distribute or publicly communicate the work, according to the following conditions:



Attribution. All CC licenses require that others who use your work in any way must give you credit the way you request, but not in a way that suggests you endorse them or their use. If they want to use your work without giving you credit or for endorsement purposes, they must get your permission first.



NonCommercial. You let others copy, distribute, display, perform, and (unless you have chosen NoDerivatives) modify and use your work for any purpose other than commercially unless they get your permission first.



NoDerivatives. You let others copy, distribute, display and perform only original copies of your work. If they want to modify your work, they must get your permission first.

There are no additional restrictions. You cannot apply legal terms or technological measures that legally restrict doing what the license allows.

This working paper was elaborated in the context of INCASI Network, a European project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie GA, No. 691004, and coordinated by Dr. Pedro López-Roldán. This article reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

Digital edition: <https://ddd.uab.cat/record/188612>

Dipòsit Digital de Documents
Bellaterra, Cerdantola del Vallès (Barcelona)
Universitat Autònoma de Barcelona



Introducción a la lógica multinivel: un análisis longitudinal con SPSS y R

Mattia Vacchiano
Jordi Merino Noé

Resumen

Este artículo propone una breve introducción al análisis multinivel. El objetivo del texto es facilitar algunos principios teóricos y metodológicos básicos para el uso de esta técnica de análisis en la investigación sociológica. La potencialidad de esta técnica se fundamenta en una lógica que permite medir la importancia del contexto en el que se generan las observaciones, midiendo la influencia e interacción entre los distintos niveles en el que se articulan los datos. En este sentido, los modelos multinivel constituyen la metodología de análisis más adecuada en el momento actual para tratar datos estructurados de forma jerárquica. El texto se divide en dos partes: una primera sección argumenta los conceptos teóricos de base de la técnica multinivel. En una segunda sección se propone un ejemplo de aplicación de un modelo de regresión logística multinivel (RLM) a un estudio sobre la inserción laboral juvenil en el Área Metropolitana de Barcelona.

Palabras clave

Análisis multinivel, SPSS, R, Inserción juvenil, Datos longitudinales, Redes sociales

Índice

1. Introducción. La lógica multinivel: la importancia del contexto. Los modelos multinivel: conceptos clave. ¿Cuándo se puede aplicar el análisis multinivel? 2. La inserción juvenil en Barcelona: una guía al análisis multinivel con SPSS y R. El modelo multinivel: estructura, objetivos e hipótesis. El proceso de modelaje en R. El proceso de modelaje en SPSS. 3. Conclusiones. 4. Referencias.

1. Introducción

1.1. La lógica multinivel: la importancia del contexto

Una de las preocupaciones de la sociología, en tanto que disciplina, ha sido intentar medir las regularidades sociales y procurar entender qué factores contextuales inciden y ayudan a mejorar la comprensión de un fenómeno. Con ello, tradicionalmente, las ciencias sociales han utilizado técnicas de análisis cuantitativas para analizar conjuntamente las características de una *unidad de análisis* (personas, organizaciones,

países) con otras variables de contexto, si bien estas últimas no se configuran únicamente como un atributo de la unidad misma. Los análisis de este tipo resultan inexactos, ya que tratan los datos como si estos se articulasen en un único nivel cayendo en la llamada «falacia ecológica»: es decir, atribuyen erróneamente que las características de un contexto inciden en todas las unidades de análisis (Murillo Torrecilla, 2008).

Para obviar este problema cabe considerar que las observaciones que investigamos se articulan en múltiples niveles. Significa que algunas

variables están agrupadas dentro otras variables y, en este sentido, los datos que estamos analizando se organizan en un orden jerárquico. Es el caso, por ejemplo, de una investigación que quiera analizar la nota media en historia del arte de un grupo de estudiantes: (i) los estudiantes están agrupados en aulas o cursos; (ii) las aulas están agrupadas en escuelas o centros educativos; (iii) las escuelas están agrupadas en distritos o provincias (Figura 1).

Aulas, escuelas o distritos no son simple atributos de los estudiantes. En una misma aula se comparten experiencias, métodos de

enseñanza y docentes que pueden determinar o influir en los resultados escolares. Es decir, estos factores contextuales pueden incidir en la variabilidad de la característica que queremos investigar: la nota en historia del arte. Con ello, ejemplificando, puede que se observe una cierta homogeneidad en los resultados en historia del arte entre los estudiantes que frecuentan la misma clase. Tener en cuenta esta homogeneidad e influencia estructural (clase, docentes, experiencias compartida) sobre ella es, en definitiva, el objetivo principal del análisis multinivel.

Figura 1. Ejemplo de datos estructurados en dos niveles. Estudiantes (nivel 1) y clases (nivel 2).

Nivel 2	clase 1	clase 2	clase 3	...	clase n
Nivel 1	estudiante 1 estudiante 2 estudiante 3 estudiante 4 estudiante 5 estudiante 6	estudiante 7 estudiante 8 estudiante 9	estudiante 10 estudiante 11 estudiante 12 estudiante 13		estudiante n

Fuente: Elaboración propia

1.2. Los modelos multinivel: conceptos clave

Si las observaciones que investigamos se articulan en múltiples niveles, entonces la estrategia de análisis multinivel nos permite trabajar con dos (o más) unidades de análisis, dependiendo del número de niveles en el que se articula la estructura jerárquica de nuestros datos. Es el investigador/a el que establece la clasificación de los niveles según sus objetivos, modelo de análisis y bagaje teórico. Con ello, en definitiva, los modelos multinivel¹ representan

ampliaciones de los modelos de regresión lineal y logística, ya que nos permiten elaborar modelos de regresión para cada nivel de análisis, permitiéndonos entender la interacción en cada uno de ellos (Murillo Torrecilla, 2008). Para comprender la aportación de una lógica de este tipo al análisis, es necesario ilustrar dos aspectos capaces de evidenciar la potencialidad de la técnica multinivel: (1) la diferencia entre *random effect* y *fixed effect* y (2) la correlación intraclase (Snijders et al., 1995):

1. Para entender la distinción entre *random effect* (coeficiente aleatorio) y *fixed effect* (coeficiente fijo), cabe considerar que en los modelos de

¹ A nivel general se habla de modelos multinivel, pero también son denominados modelos jerárquicos (por la estructuración de los datos), modelos mixtos o mixtos categóricos (porque utilizan parámetros fijos y

parámetros aleatorios) y también de regresión lineal o logística multinivel (Merino, 2017).

regresión lineal clásicos los parámetros, el intercepto y la pendiente, son comunes a todos los sujetos: es decir, son estimados a partir de todas las observaciones y se asumen como coeficientes fijos, representativo de toda la muestra. En una estructura multinivel, en cambio, estos coeficientes son aleatorios: es decir, cada grupo tiene su propio intercepto y pendiente, como representación específica de cada contexto. Con ello, los coeficientes de regresión del nivel “uno” se regresan en el nivel superior, y así sucesivamente hasta obtener un modelo general. Esta distinción representa la gran aportación del análisis multinivel, ya que la composición de los modelos multinivel en dos partes (una parte general, fija, y una variable, que varía en cada grupo) implica que las variables interactúan en distintos niveles (*interacción internivel*). Así pues, básicamente se puede estimar el efecto de las variables — ó factores — explicativas del fenómeno a través de los efectos fijos y la estimación de qué parte de la variabilidad podría ser imputable a cada nivel a través de los efectos aleatorios.

2. La *correlación intraclase* es la medida del grado de dependencia de los individuos (Murillo Torrecilla, 2008). Es una estimación que interpreta qué similares pueden ser las unidades de análisis en un mismo grupo y, en este sentido, nos ayuda a comprender cuanto los resultados de nuestro modelo general podrían depender de la homogeneidad del contexto. Si el valor de esta medida se acerca a cero, significa que los grupos no son homogéneos internamente y por lo tanto este nivel grupal (este contexto) no influye en los resultados, ya que ser o no ser parte de un grupo es prácticamente irrelevante en el análisis. Es decir, si pertenecer a un grupo o contexto es importante y hay puntuaciones en los datos bastante similares a nivel interno la diferencia entre grupos será más acentuada al existir una mayor variabilidad entre ellos, por tanto, un valor mayor. Este valor ofrecido por el coeficiente de correlación intraclase (ICC) es uno de los aspectos más importantes de tener en cuenta a la hora de

evaluar un modelo ya que facilita una medida de la similitud entre los grupos, indicando qué parte de la variabilidad total es debida a la variabilidad de las observaciones en cada nivel².

Tabla 1. Comparación entre probabilidad, odds y log odds.

Probabilidad	odds	log odds
.001	.0010010	-6.906755
.01	.0101010	-4.595120
.15	.1764706	-1.734601
.2	.25	-1.386294
.25	.3333333	-1.098612
.30	.4285714	-.8472978
.35	.5384616	-.6190392
.4	.6666667	-.4054651
.45	.8181818	-.2006707
.5	1	0
.55	1.222222	.2006707
.6	1.5	.4054651
.65	1.857143	.6190392
.7	2.333333	.8472978
.75	3	1.098612
.8	4	1.386294
.85	5.666667	1.734601
.999	999	6.906755
.9999	9999	9.210240

Fuente: Fuente (última consulta Marzo 2017): Institute for Digital Research (<http://stats.idre.ucla.edu/>).

Aunque la intención inicial de los modelos multinivel fue la de ampliar los modelos de regresión lineal, los investigadores/as han ido extendiendo la lógica multinivel también a los estudios con variables dependientes de tipo categórico desde hace algunos años (Heck *et al.*, 2012). En este caso, siendo la variable respuesta de carácter dicotómico, se asume que su transformación logarítmica realiza una relación lineal con los factores predictores. Con ello, los resultados de los modelos de RLM se basan en coeficientes de regresión, cuya interpretación se basa en los conceptos de *probabilidad*, *odds ratio* (u Oportunidades Relativas Asociadas, ORA) y

² El coeficiente de correlación intraclase (ICC) se aplica principalmente en los modelos de regresión lineal multinivel, puesto que su estimación proviene del análisis de la varianza residual. En este sentido, la capacidad del ICC de medir la homogeneidad interna de los grupos o niveles en modelos con variables categóricas (esto es, en los modelos de regresión logística multinivel) es objeto de debate (Bauer y Sterba, 2011; Berkhof y Snijders, 2001)

log odds. Muy brevemente, asumido el concepto de probabilidad como un valor que varía entre 0 y 1, el odds ratio se define como la posibilidad de que un evento ocurra frente a la posibilidad de que no ocurra (Szumilas, 2010), esto es, la posibilidad que nuestra variable respuesta asuma un valor en base a la referencia. La transformación desde la probabilidad al odds implica que al aumentar la probabilidad aumenta también el valor del odds. Con ello, los resultados de los modelos RLM pueden expresarse también mediante el logaritmo del odds (log odds), el cual puede asumir la forma negativa o positiva (Tabla 1), o también a partir de la interpretación exponencial de éste con un intervalo de confianza para estudios que buscan proporciones de que el fenómeno suceda.

1.3. ¿Cuándo se puede aplicar el análisis multinivel?

Como se ha mencionado, para realizar un análisis multinivel cabe considerar y definir la estructura jerárquica en la que se organizan nuestros datos. En primer lugar, cuando las observaciones se generan en un determinado contexto (o nivel) es necesario tener en cuenta el problema de la dependencia entre las observaciones. Significa que el comportamiento de una determinada observación puede influir o depender del comportamiento de otra, aspecto que puede observarse precisamente entre observaciones que se generan en el mismo contexto.

El análisis multinivel, entonces, se aplica para todos aquellos conjuntos de datos en el que las observaciones no son independientes, y que además puedan estar agrupadas en diferentes niveles. Esta atención hacía el tipo de estructuración de los datos soluciona el problema de la dependencia entre los datos, y permite obtener resultados más robustos y fiables (Cebolla, 2013). El análisis multinivel, por tanto, se puede utilizar en todos aquellos estudios en el que:

1. Los datos se estructuren de forma jerárquica.
2. Se quiera trabajar en un mismo modelo diferentes predictores tanto individuales

como contextuales sobre una variable dependiente (*interacción intranivel*).

3. Se quiera observar por separado la incidencia y varianza dentro del grupo con los otros niveles de agrupación (*correlación intraclase*).

2. La inserción juvenil en Barcelona: una guía al análisis multinivel con SPSS y R

En esta segunda sección proponemos un ejemplo de aplicación de un modelo de regresión logística multinivel en un estudio sobre la inserción laboral juvenil de 250 jóvenes entre 20 y 34 años en el Área Metropolitana de Barcelona. El estudio, denominado «Redemas»³ y realizado por el Centre d'Estudis Sociològics sobre la Vida Quotidiana i el Treball (QUIT) ha sido financiado por el Ministerio de Economía y Competitividad (Ref: CSO2012-36055) con el objetivo, entre los demás aspectos, de comprender el peso de los contactos personales como vía de acceso al empleo antes y durante la crisis económica. Con ello, el estudio obtuvo información socioeconómica sobre la persona encuestada, reconstruyó las trayectorias laborales de los entrevistados/as de manera retrospectiva y obtuvo, mediante un generador de nombres, el listado de 20 personas que componen la red social de los encuestados/as, analizando el peso de estos contactos como vía de acceso al empleo.

Los datos longitudinales obtenidos para reconstruir las trayectorias laborales –que utilizamos para construir nuestro modelo multinivel– se basan en una *life history grid*, esto es, una parrilla temporal en que se van situando todas las actividades laborales (y formativas) que componen la trayectoria de la persona encuestada. De esta manera, la muestra de datos recoge los eventos educativos o de empleo, los acontecimientos laborales de carácter informal y los períodos de desempleo, con o sin prestación. Asimismo, para cada actividad laboral (además de conocer sus

³ El nombre completo del proyecto es: «Las redes sociales, en sus diferentes modalidades, como recursos y mecanismos de búsqueda laboral e inserción laboral en el empleo y de apoyo social en los jóvenes».

características), la muestra recoge el mecanismo de acceso al empleo (por vía formal o por contacto personal).

2.1. El modelo multinivel: estructura, objetivos e hipótesis

En el marco de este estudio, por tanto, aplicamos la lógica multinivel para comprender los factores explicativos de la inserción laboral juvenil mediante las redes personales. Precisamente, nos preguntamos si la obtención de empleos (tanto formales como informales) mediante contactos puede explicarse por alguna característica de los jóvenes entrevistados/as, o bien por las características de los empleos obtenidos o, diversamente por qué estas inserciones laborales se han producido durante o antes de la crisis económica⁴.

Para responder a esta pregunta, considerada la estructura de los datos del proyecto Redemas, necesitamos implementar un modelo de regresión logística multinivel, como especificamos más adelante⁵.

a. Organización de los datos

Como primer paso, cabe identificar la estructura jerárquica en el que se organizan los datos de la matriz Redemas.sav. Nótese que en las primeras dos columnas se señalan los niveles que componen la matriz: las filas representan los empleos (o eventos laborales) que componen la carrera laboral, y por lo tanto representan el nivel más bajo de la estructura (ID_evento_Nivel1). En la primera columna, en cambio, la variable (ID_caso_Nivel2) representa el número de identificación de los 250 jóvenes de la muestra, lo cual nos indica la

trayectoria laboral de pertenencia de los empleos. En este caso consideramos:

- los empleos (eventos laborales) como nuestra unidad de análisis (nivel 1)
- los jóvenes como el “contexto” en el que se generan las observaciones (nivel 2)

¿Por qué aplicamos la lógica multinivel? Porque nuestras observaciones o unidades de análisis (los empleos) no son independientes. Precisamente, los empleos, como primer nivel en la estructuración de nuestros datos, se generan en el contexto de la trayectoria laboral de un joven, que por tanto representa el segundo nivel de nuestros datos. Por lo tanto, cabe considerar como la variabilidad de las características de los empleos que componen una trayectoria es dependiente de las características del joven, que en este caso representaría nuestro contexto.

En definitiva, aplicamos un análisis multinivel ya que:

1. Los casos no son independientes y se estructuran de forma jerárquica o anidada. Los eventos laborales tienen algo en común, la pertenencia a la trayectoria laboral de un joven.
2. Se quiere trabajar en un mismo modelo diferentes predictores, relacionados tanto con las características de los empleos como con las características de los jóvenes sobre una variable dependiente (interacción internivel).
3. Se quiere observar por separado la incidencia y varianza de los eventos laborales, a nivel grupal, como al nivel de los jóvenes (correlación intraclase).

b. Objetivos, hipótesis y modelo

El planteamiento de un modelo multinivel, como toda técnica de análisis, responde de los objetivos de investigación y las hipótesis específicas relacionadas con el objeto de estudio. En este caso hemos decidido centrar nuestro interés en una variable: el canal de acceso al empleo (V7.Metodo.de.inserción en la matriz Redemas.sav).

Esta variable nos indica como ha sido

⁴La inserción mediante redes personales ha sido asociada tanto a los ciclos económicos (Requena, 1991) como a la obtención de empleos poco cualificados (Rieucan, 2008) o a las características de las personas que buscan empleo, especialmente aquellos pertenecientes a grupos más vulnerables (Smith, 2000).

⁵ Es posible acceder a la matriz de datos **Redemas.sav** a través del siguiente enlace: <http://incasi.uab.cat/sites/incasi.uab.cat/files/Redemas.sav>.

obtenido el empleo, distinguiendo entre: mecanismos formales (anuncios, convocatorias u ofertas públicas de empleo) o mecanismos informales (contactos personales). Nuestra hipótesis de partida es que las redes, por sus características de informalidad y rapidez, se han convertido en un canal de acceso al empleo prioritario en tiempos de crisis. En este sentido, planteamos un modelo de regresión logística multinivel para medir si los eventos laborales obtenidos a través los contactos personales han aumentado en los años de recesión económica. Además, también esperamos que los empleos obtenidos a través

los contactos se asocien con mayor probabilidad a empleos precarios, irregulares o complementarios.

A partir de estas hipótesis, planteamos un modelo de regresión logística binomial multinivel. La variable dependiente (VD) es el mecanismo de acceso al empleo (formales o informales) mientras las variables independientes (VI) están relacionadas con las características del evento y las características de los y las jóvenes. En este caso, incluimos en nuestro modelo cuatro variables dicotómicas (Tabla 2).

Tabla 2. Definición de variables

Variables	Nivel	Definición	
			<i>Dependiente</i>
Mecanismo formal o informal de acceso al empleo <i>V7.Metodo.de.inserción</i>	uno	El empleo ha sido obtenido por contacto personal [1] o por otros mecanismos de inserción [0]	
			<i>Independientes</i>
Importancia de la actividad <i>V5.Importancia.actividad.</i>	uno	Indica si el empleo es una actividad laboral principal [1] o una actividad secundaria y complementaria [0]	
Relación laboral <i>V9.1. Tipodecontrato_illegal.</i>	uno	Identifica la relación laboral del empleo: con contrato [0] o sin contrato [1]	
Inserción en contexto de crisis <i>V10.Crisis.</i>	uno	Señala si el empleo se ha obtenido antes de la crisis (1996 – 2008) [0] o durante la crisis (2008-2014) [1]	
Sexo del entrevistado/a <i>V13.Sexo.</i>	dos	El sexo del entrevistado, mujer [0] o hombre [1]	

2.2. El proceso de modelaje en R

a. Instalación de RStudio

Presentamos el procedimiento para construir el modelo multinivel con R. Sabemos que este programa es un lenguaje de programación orientado a objetos y que es un software libre y abierto, lo cual representa una primera ventaja en comparación al uso del programa SPSS.

En primer lugar, necesitamos instalar el programa RStudio a través: <http://rstudio.org/>. Este programa permite la visualización simultánea de los entornos de trabajo de R, y mejora la eficiencia y usabilidad del programa.

RStudio se organiza en cuatro áreas de trabajo

diferenciadas⁶:

1. En el área en alto a la izquierda podemos abrir y editar ficheros con código R, denominados *Scripts*. Este código representa el lenguaje de R.
2. En la parte inferior izquierda podemos observar la *console* de R, a través la cual ejecutamos las operaciones en R y visualizamos los resultados.
3. En alto a la derecha encontramos las ventanas del *History* (el historial de las operaciones en R, con las líneas de código ejecutada) y *Environment* (donde visualizamos el listado de objetos creados en R).

⁶ Esta información se ha extraído del material de la asignatura “Mètodes Quantitatius de Recerca en Criminologia” del Dr. Pedro López-Roldán y Dra. Sandra Fachelli.

4. Bajo a la derecha, finalmente, encontramos cuatro ventanas:
- *Files*: para acceder a los ficheros del disco duro;
 - *Plots*: para visualizar los gráficos;
 - *Packages*: para la gestión de los paquetes instalados.
 - *Help*: para acceder a la ayuda en línea.

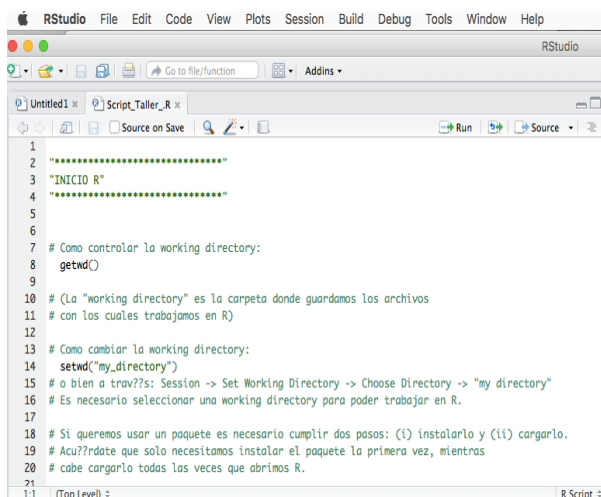
b. Ejecutar el modelo

Procedemos paso a paso para ejecutar el modelo. En primer lugar, cabe controlar nuestra *working directory*. La *working directory* es la carpeta donde guardamos todos los archivos con los cuales trabajamos en R. Para saber cuál es nuestra *working directory* ejecutamos el comando: `getwd()` en la consola de R (ventana baja a la izquierda).

Con todo, podemos elegir la carpeta donde trabajar a través el comando `setwd()` (“Nombre carpeta” o bien a través el Menú, a través *Session* → *Set Working directory* → *Choose Directory...*

Elegiremos la carpeta donde tenemos guardado nuestra matriz de datos *Redemas.sav*.

Figura 2. Visualización del código del fichero *Script.R*



Una vez seleccionada la *working directory*, podemos guardar en la carpeta también el file

Script.R.⁷

Sucesivamente, será posible visualizar el contenido de este texto a través el menú contextual *File* → *Open File...*

Después de haber seleccionado nuestra *working directory*, si queremos ejecutar una función es necesario cumplir dos pasos: (i) instalar los paquetes necesarios y (ii) cargarlos. En el lenguaje de R, un paquete es una recolección de funciones, datos y códigos que nos permiten operar en el entorno de R. Solo necesitamos instalar el paquete la primera vez, mientras cabe cargarlo todas las veces que abrimos R.

Por lo tanto, el paso siguiente es instalar los paquetes necesarios para poder ejecutar la función que nos permite calcular nuestro modelo multinivel. Con ello, tenemos que ejecutar en la consola el siguiente código:

```

install.packages("lme4")
install.packages("optimx")
install.packages("RVAideMemoire")
install.packages("sjPlot")
install.packages("arm")
install.packages("VGAM")
install.packages("MASS")

```

Sucesivamente, el comando “*Library*” nos permite cargar los paquetes:

```

library(foreign)
library(lme4)
library(optimx)
library(RVAideMemoire)
library(sjPlot)
library(MASS)
library(arm)
library(VGAM)

```

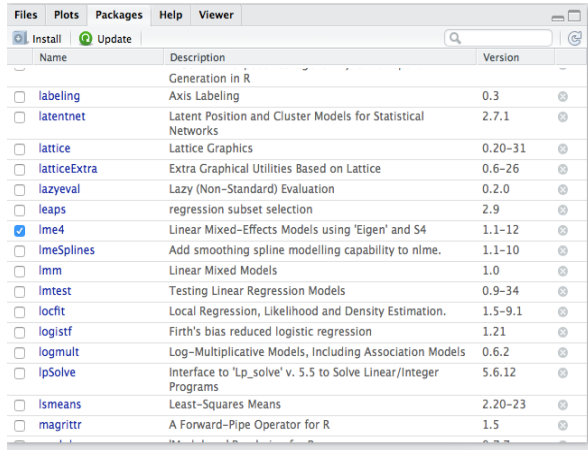
Podemos comprobar que los paquetes hayan sido instalados y cargados correctamente a través la ventana “*Packages*” en la parte baja a la derecha de *Rstudio* (Figura 3).

⁷ Mediante el enlace a la página <http://incasi.uab.cat/sites/incasi.uab.cat/files/Script.R> es posible acceder también al file *Script.R* para visualizar todos los comandos necesarios a la realización del modelo presentado en este texto.

Una vez instalados y cargados los paquetes, el paso siguiente es cargar en R nuestra matriz de datos, a través los comandos:

```
Matriz <- read.spss("Redemas.sav", to.data.frame=TRUE)
attach (Matriz)
```

Figura 3. La ventana “Packages”

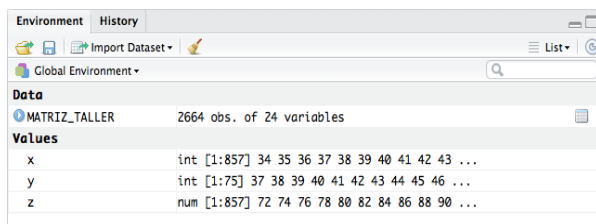


Una vez instalados y cargados los paquetes, el paso siguiente es cargar en R nuestra matriz de datos, a través los comandos:

```
Matriz <- read.spss("Redemas.sav", to.data.frame=TRUE)
attach (Matriz)
```

Podemos averiguar que la matriz ha sido cargada correctamente a través la ventana “Environment” (Figura 4).

Figura 4. La ventana “environment”⁸

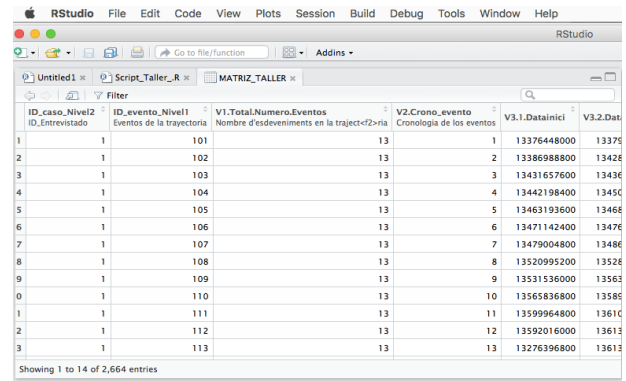


Visualizamos la matriz en la ventana en alto a la derecha clicando dos veces sobre el objeto *Matriz* a través la misma ventana (Figura 5).

Una vez cumplidos todos estos pasos podemos efectivamente ejecutar el código de la función *glmer* (*Generalized Linear Mixed-Effects Models*).

⁸ En la visualización el objeto *Matriz* se nombra *Matriz_taller*

Figura 5. Visualización de la matriz de datos



La fórmula separa la variable dependiente, a la izquierda, con el operador \sim y las variables independientes con el operador $+$, a la derecha. Los efectos aleatorios se distinguen a través la barra vertical ($|$).

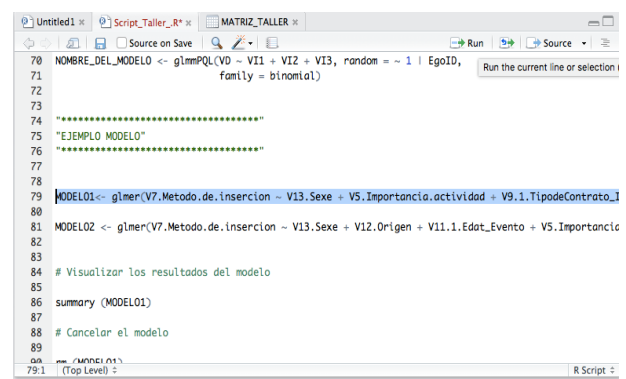
```
MODELO <- glmer(VD ~ VI + VI + VI + VI + (1
| "VARIABLE_NIVEL2"), family = binomial(logit),
control=glmerControl(optimizer="bobyqa", optCtrl =
list(maxfun = 100000)))
```

Creamos el objeto “MODELO1” e insertamos las variables del modelo en las partes correspondientes:

```
MODELO1<- glmer(V7.Metodo.de.insercion ~ V13.Sexe +
V5.Importancia.actividad + V9.1.TipodeContrato_Illegal +
V10.Crisis + (1 | ID_caso_Nivel2), family = binomial(logit),
control=glmerControl(optimizer="bobyqa", optCtrl = list(maxfun
= 100000)))
```

Es importante saber que todos estos pasos pueden ser ejecutados directamente a través el código presente en el *Script*. En este caso podemos subrayar la línea de código del Modelo y clicar el botón RUN, en alto a la derecha de la ventana.

Figura 6. La línea de código del modelo



Una vez ejecutado este comando notaremos haber creado el objeto “MODELO1” en la ventana “*Environment*”. Significa que ya disponemos de nuestros resultados y solo necesitamos visualizarlo a través el comando:

`summary(MODELO1)`

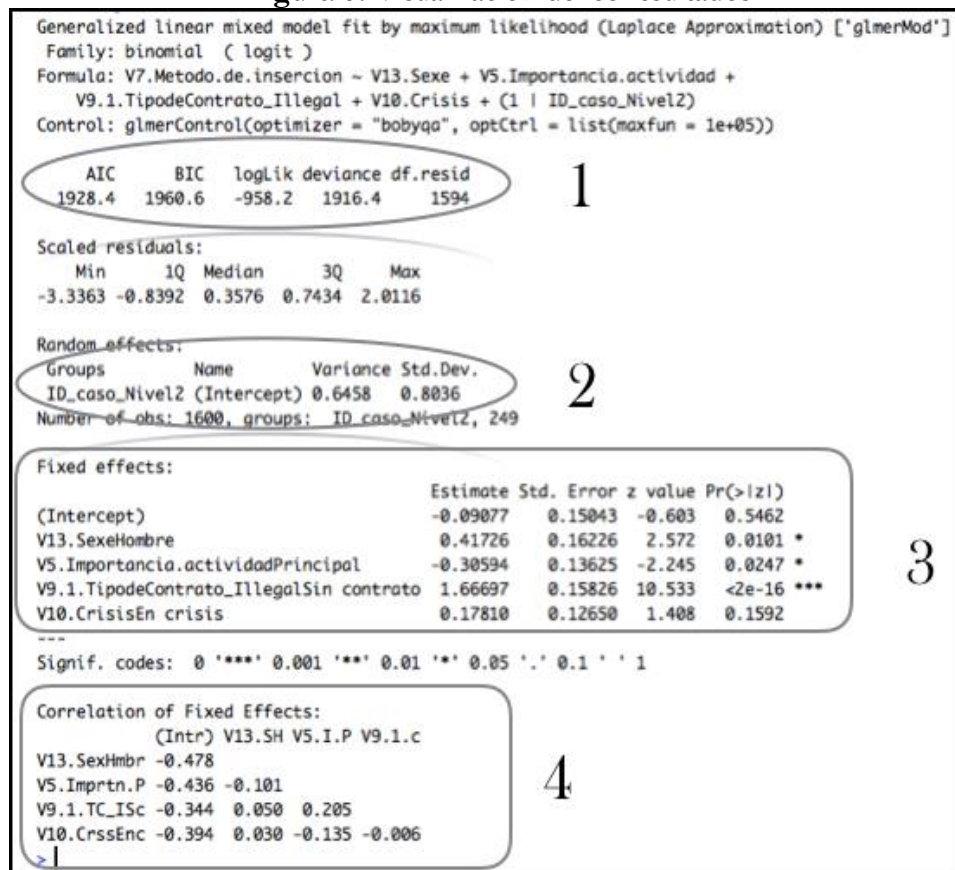
Comentamos punto por punto los resultados obtenidos:

1. Como en una regresión logística para conocer la bondad de ajuste disponemos de distintos test. Dos son los principales:

- Akaike’s information criterion (AIC)
- Schwarz’s Bayesian criterion (BIC)

Ninguno de los dos test puede ser intrínsecamente interpretable. Esencialmente valores más bajos señalan una mayor bondad de ajuste, aunque algunos autores señalan que estos indicadores son más adecuados en el caso de regresiones lineales donde tiene más sentido apreciar las sensibilidades en los procesos secuenciales a medida que se incorporan nuevas variables o covariables en el modelo (Bauer y Sterba, 2011).

Figura 7. Visualización de los resultados



2. Como sabemos los modelos multinivel se componen de dos partes: una parte fija y una variable. El apartado “*Random effect*” señala la desviación estándar de la parte variable, permitiéndonos calcular el coeficiente de correlación intraclase (ICC), es decir el porcentaje (%) de varianza de la VD explicada por el factor aleatorio (en este caso, los jóvenes). Para calcular el ICC usamos la

formula:

$$(STdev)^2 / (3.29 + (STdev)^2).$$

En nuestro modelo, esto es: $ICC = 0.6458 / (3.29 + 0.6458) = 0.164$. Esto se puede interpretar de forma porcentual indicando que el 16,4% de la probabilidad que un empleo se consiga por contacto personal depende de las características del joven.

3. En este apartado visualizamos efectivamente nuestro modelo general. En el modelo se evidencian:

- *Estimate*: los coeficientes estimados por cada variable.
- *St. Error*: la desviación estándar
- *Z value*: los residuos
- *P value*: la significación

Notamos que nuestra hipótesis se confirma solo parcialmente. Los coeficientes señalan una asociación positiva de las variables “sexo”, “tipo de contrato” y “crisis” (aunque solo las primeras dos son significativas tal y como se aprecia en el valor *Pr* donde están por debajo de 0,05) y una asociación negativa en la variable “importancia de la actividad”. En este caso:

- el modelo destaca una mayor probabilidad que los empleos obtenidos por contacto personal sean trabajos sin contrato, obtenido por hombres, y que no sean una actividad principal, sino secundaria o complementar a otra actividad de estudio o de trabajo, como planteado en nuestra hipótesis.
- Sin embargo, la variable crisis, si bien se asocia positivamente, no es significativa, por lo tanto, no podemos afirmar que obtener empleos por contacto personal sea más probable en una época de crisis, sino más bien este mecanismo de acceso al empleo se configura como una característica estructural del mercado laboral en el contexto investigado.

4. Finalmente, en el último apartado visualizamos las correlaciones entre las variables. Esta matriz nos permite identificar problemas de multicolinealidad en el modelo general. En nuestro modelo no se observan.

2.3. El proceso de modelaje en SPSS

El procesador IBM SPSS es un paquete estadístico históricamente utilizado en las universidades, en combinación con otros paquetes, y con una implantación considerable en el mercado privado. A partir de su versión 19.2 incorpora la posibilidad de poder realizar

modelos multinivel con variables respuesta de tipo categórico a través de los denominados modelos lineales mixtos generalizados (GLMM). Con las versiones posteriores se han ido mejorando los procesos y el manejo de diferentes tipos de variables y funciones, como es el caso de la versión 22.0.

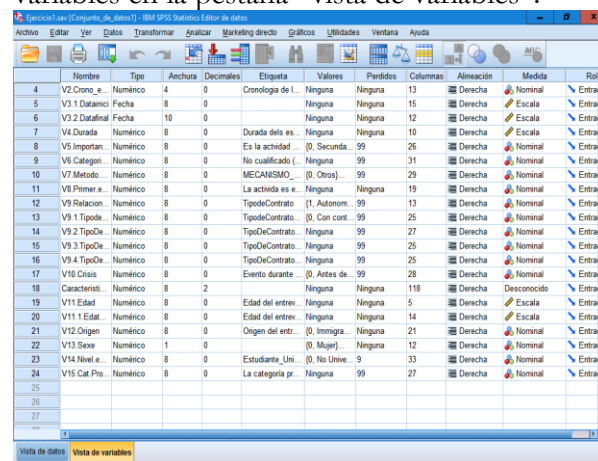
Es central que antes de proceder a la explotación estadística el investigador/a tenga clara la estructura del modelo que desea realizar. En primer lugar, hay que tener en cuenta que la composición de las variables disponga de la estructura necesaria para proceder a su análisis. La variable respuesta ha de ser dicotómica y las variables predictoras han de ser categóricas (nominales u ordinales) o también variables de tipo lineal.

Este modelo responderá a la pregunta de investigación que el investigador/a quiera hacer y ha de estar sustentado en algún bagaje teórico que facilite la posterior interpretación de los resultados. Siguiendo el modelo anterior realizaremos el modelo paso a paso con SPSS.

a. Ejecución del modelo

Los datos se estructuran en diferentes ventanas: una de visualización de la información (dónde se puede apreciar la matriz de datos y el listado de variables), una ventana de sintaxis (que recoge los procedimientos utilizados) y una ventana de resultados (ver Figura 8).

Figura 8. Distribución de la información de las variables en la pestaña “vista de variables”.



Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Almacenamiento	Medida	Rol
V2 Crono_e	Número	4	0	Cronología de l.	Ninguna	Ninguna	13	Derecha	Nominal	Entrada
V3 1 Datafnci	Fecha	8	0		Ninguna	Ninguna	15	Derecha	Escala	Entrada
V3 2 Datafnci	Fecha	10	0		Ninguna	Ninguna	12	Derecha	Escala	Entrada
V4 Durada	Número	8	0	Durada de la es.	Ninguna	Ninguna	10	Derecha	Escala	Entrada
V5 Importan.	Número	8	0	Es la actividad	(0, Secunda	99	26	Derecha	Nominal	Entrada
V6 Categ.	Número	8	0	No cualificado	Ninguna	99	31	Derecha	Nominal	Entrada
V7 Metodo.	Número	8	0	MECANISMO	(0, Otros)	99	29	Derecha	Nominal	Entrada
V8 Primer e.	Número	8	0	La actividad es e.	Ninguna	Ninguna	19	Derecha	Nominal	Entrada
V9 Relación.	Número	8	0	TipoDeContrato	(1, Autonom	99	13	Derecha	Nominal	Entrada
V9 1 Tipode.	Número	8	0	TipoDeContrato	(0, Con cont	99	25	Derecha	Nominal	Entrada
V9 2 Tipode.	Número	8	0	TipoDeContrato	Ninguna	99	27	Derecha	Nominal	Entrada
V9 3 Tipode.	Número	8	0	TipoDeContrato	Ninguna	99	25	Derecha	Nominal	Entrada
V9 4 Tipode.	Número	8	0	TipoDeContrato	Ninguna	99	25	Derecha	Nominal	Entrada
V10 Crisis	Número	8	0	Evento durante	(0, Antes de	99	28	Derecha	Nominal	Entrada
V11 Caracteris.	Número	8	2		Ninguna	Ninguna	118	Derecha	Desconocido	Entrada
V11 Edad	Número	8	0	Edad del entrevist.	Ninguna	Ninguna	5	Derecha	Escala	Entrada
V11 1 Edad.	Número	8	0	Edad del entrevist.	Ninguna	Ninguna	14	Derecha	Escala	Entrada
V12 Origen	Número	8	0	Origen del entr.	(0, Inmigr.	Ninguna	21	Derecha	Nominal	Entrada
V13 Sexo	Número	1	0		(0, Mujer)	Ninguna	12	Derecha	Nominal	Entrada
V14 Nivel e.	Número	8	0	Estudiante Uni.	(0, No Univ	9	33	Derecha	Nominal	Entrada
V15 Cat Pro.	Número	8	0	La categoría pr.	Ninguna	99	27	Derecha	Nominal	Entrada

Realizamos las operaciones a partir de la secuencia de comandos de la barra superior (o bien a través la ventana de *sintaxy*). Existen diferentes rutas para realizar modelos multinivel, pero en el caso de variables dependientes categóricas (dicotómicas) la ruta es la siguiente:

Análizar > Modelos Mixtos > Lineales generalizados (modelos mixtos según versión)

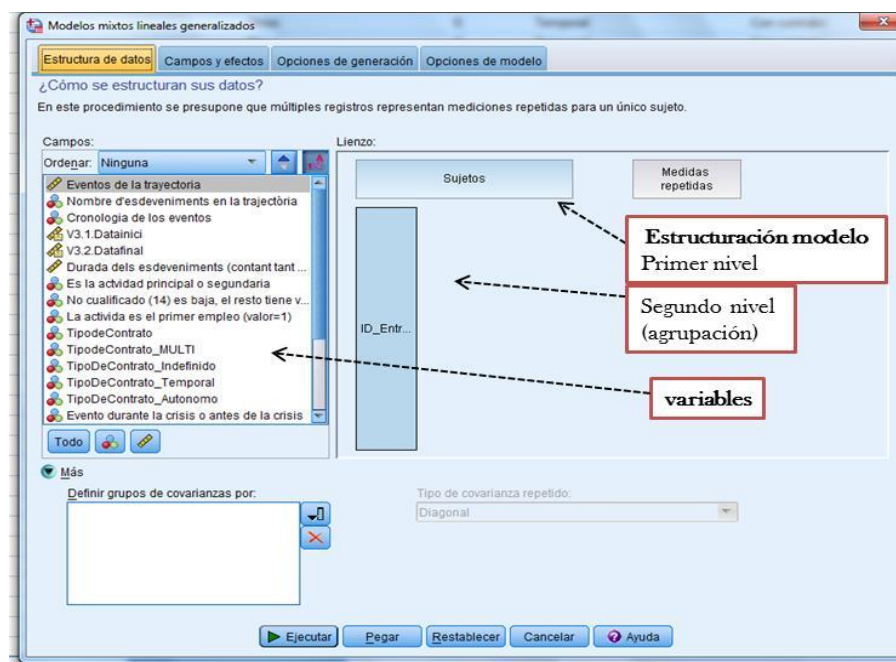
Realizamos las operaciones a partir de la secuencia de comandos de la barra superior (o bien a través la ventana de *sintaxy*). Existen diferentes rutas para realizar modelos multinivel, pero en el caso de variables dependientes categóricas (dicotómicas) la ruta es la siguiente:

Análizar > Modelos Mixtos > Lineales generalizados (modelos mixtos según versión)

Si es la primera vez que se realiza el procedimiento en un procesador SPSS es posible que aparezca una ventana de diálogo para recordar que el nivel de medición cabe realizarlo manualmente. En este caso se aconseja indicar “asignar manualmente”. Después aparecerá una ventana de diálogos donde hay diferentes pestañas para proceder al análisis paso a paso. En los cuadros de las pestañas hay diferentes opciones, pero solo se hará mención a las que tienen a ver con el modelo que se está realizando.

La primera pestaña es la de estructura de datos. En el cuadro hay que arrastrar las variables en los niveles que se quiera agrupar. Se entiende que el primer nivel lo conforman los empleos por tanto solo hay que incorporar el segundo nivel de nuestros datos, en este caso el número de identificación los jóvenes (ID_Entrevistado).

Figura 9. Cuadro de diálogo de estructuración de los datos.



En la segunda pestaña se definen los campos y los efectos (Figura 10). Estos están compuestos por cuatro acciones que aparecen en el cuadro de diálogo de la banda izquierda. En la primera acción (destino) cabe definir la variable independiente y remarcar que la regresión que se ejecutará será logística binaria. La segunda acción (efectos fijos) será indicar que variables son los predictores en el modelo.

Se aconseja no añadir ninguna variable para obtener el modelo nulo si se deseara hacer un análisis paso a paso. El modelo nulo indicaría la variabilidad del contexto sin los predictores. Para modelos más complejos, con agrupaciones de factores y covariables (variable cualitativa y cuantitativa, respectivamente), en este apartado se podrían ir incorporando según los pasos que se quiera seguir.

Figura 10. Cuadro de diálogo de diseño de las variables independientes en los efectos fijos.

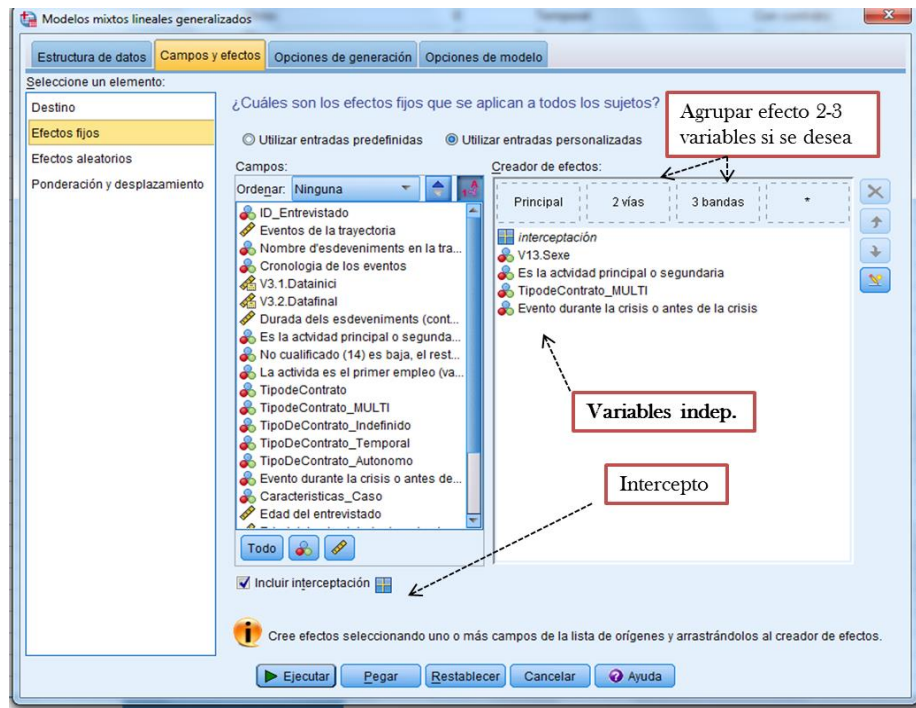
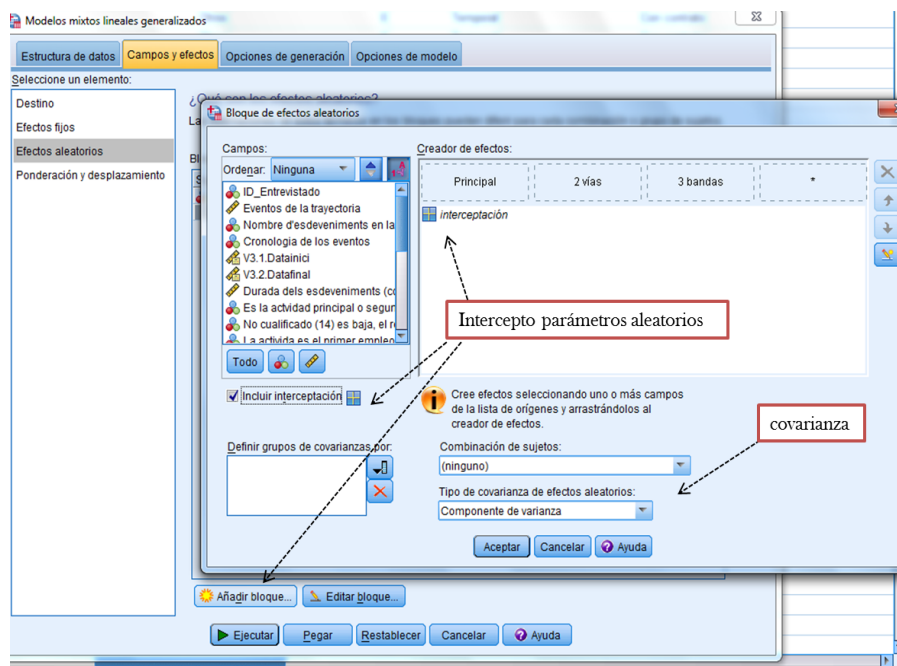


Figura 11. Cuadro de diálogo de efectos aleatorios.



La acción tercera (Figura 11) corresponde a la elección del intercepto del modelo en los parámetros aleatorios (se aconseja en el tipo de componentes de la varianza en relación a los efectos aleatorios, aunque hay otras opciones). Para modelos más compuestos se podrían incorporar otros bloques de efectos aleatorios, pero en el caso de modelos dicotómicos es

suficiente el intercepto por su tipo de distribución en dos valores.

La cuarta opción es por si hubiera que ponderar la muestra, elección que también puede hacerse mediante otros comandos de SPSS. Las opciones de generación, que son la última pestaña, indican cómo se desea ver los

resultados. Es importante remarcar el intervalo de confianza al 95% (que nos ayudará, por ejemplo, a interpretar las Oportunidades Relativas Asociadas, ORA) y la composición de

los resultados en función de la categoría de referencia que deseemos comparar, por ejemplo, hombres = 1, mujeres = 0 (ver Figura 12).

Figura 12. Opciones de generación de los resultados.

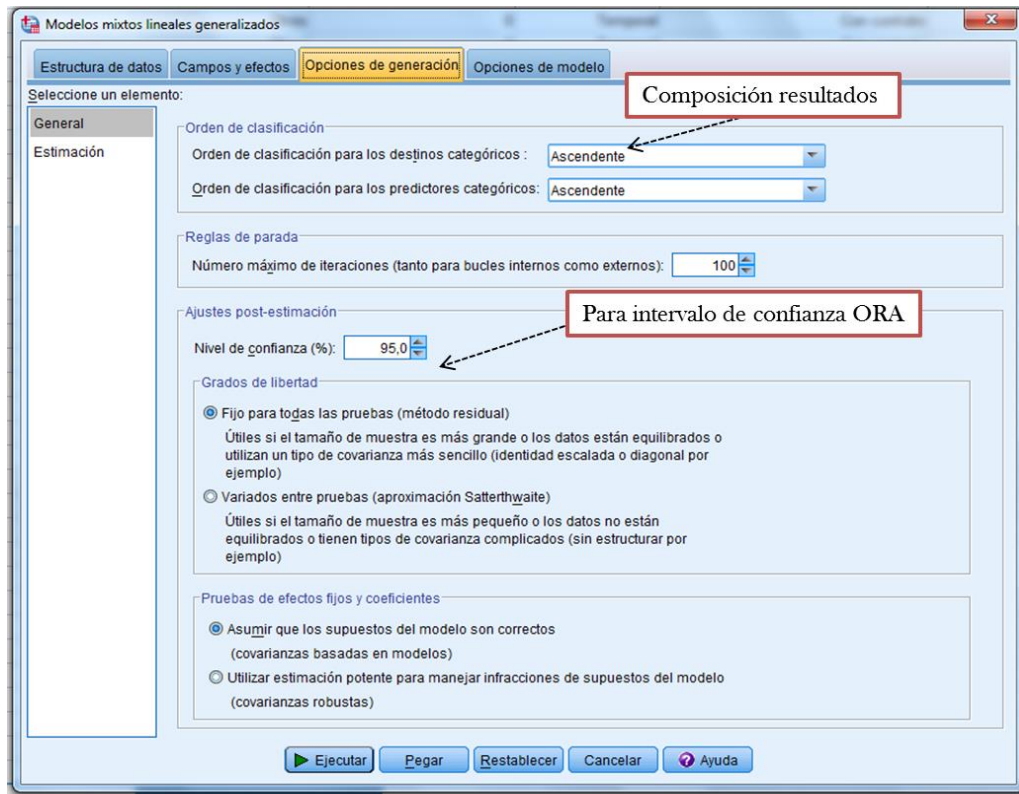
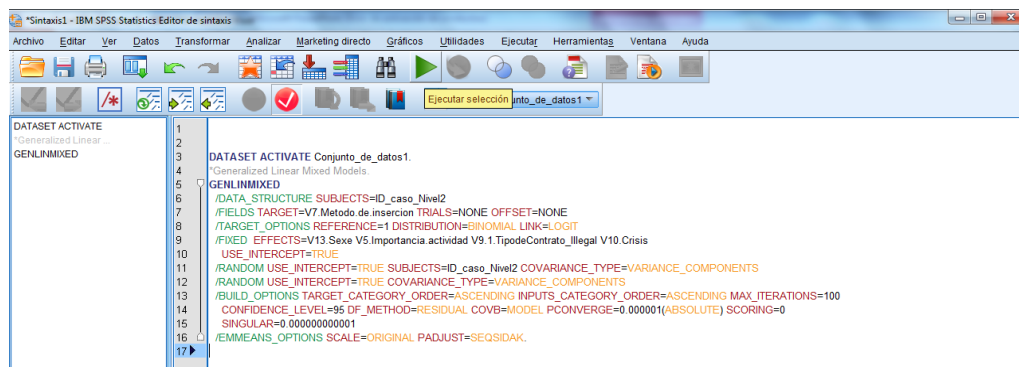


Figura 13. Ventana de *Sintaxy*.



Una vez realizado el modelo se puede proceder a indicar “ejecutar” para realizar los resultados o “pegar” para trasladarlo a lenguaje de *sintaxy* (ver Figura 13).

b. Visor de resultados.

Una vez ejecutado el análisis en la ventana de resultados aparece la información genérica del modelo y de los casos incorporados. Hay que

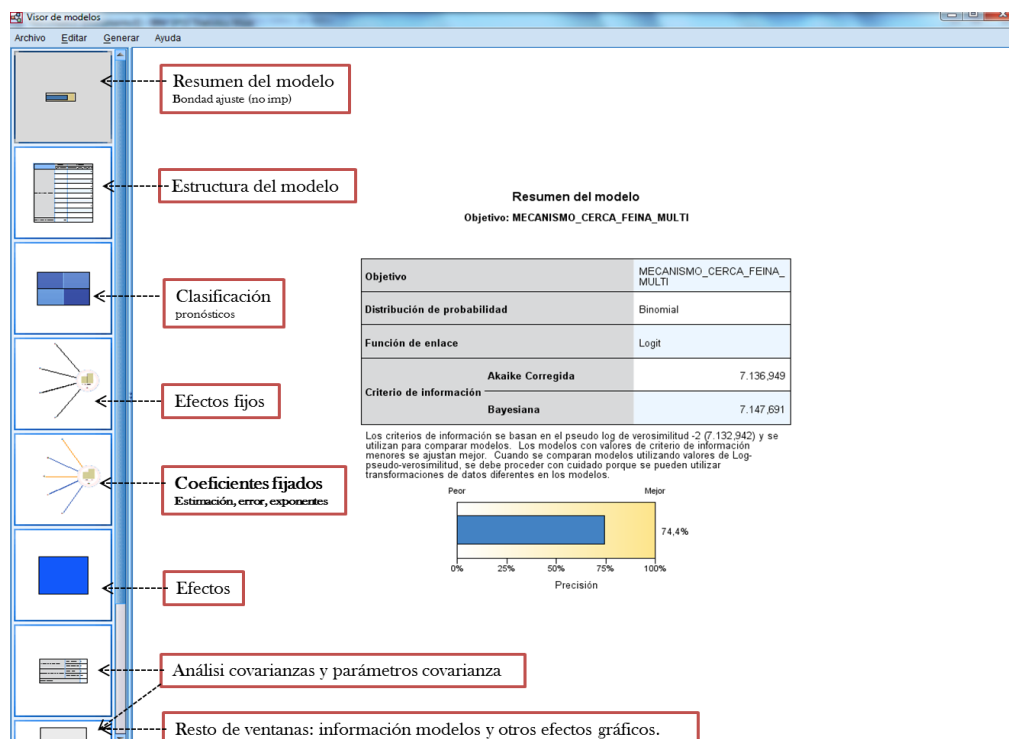
clickar dos veces en el cuadro de resultados que aparece.

En la parte izquierda del visor aparece el diálogo estructurado por parámetros (Figura 14), y permite ir directamente a la consulta de los datos. Para las primeras veces se aconseja ir a la última ventana dónde hay un resumen de todos los pasos que se han realizado en el

modelo y las opciones de construcción, así como las variables que se han utilizado, además de una información básica sobre el método y

los modelos mixtos (no se ofrece figura). Esto ayudará a ir familiarizándose y ayudar a ser más ejecutivo en futuros análisis.

Figura 14. Clasificación del visor de resultados según parámetros.



En la ventana de resumen del modelo se facilita poder consultar la bondad de ajuste del modelo a través de los criterios de información, pero como ya se ha expuesto anteriormente en los casos de RLM no son muy procedentes.

En la ventana de los coeficientes fijados la visualización de datos aparece de forma gráfica y coloreada según la asociación sea positiva o negativa y con un grosor proporcional a su peso sobre la variable dependiente. Poniendo el cursor encima de las líneas aparecerá la información sobre dicho valor. Además, en la parte inferior aparece una opción para visualizar la información del modelo mediante una tabla (Figura 15).

A través de la tabla de resultados (Figura 16) se pueden interpretar las estimaciones o las Oportunidades Relativas Asociadas (ORA), que son la elevación exponencial de las estimaciones. A nivel inicial tanto el sexo como la actividad principal o el tipo de contrato se

asocian con la inserción por redes personales, siendo esta relación estadísticamente significativa:

- Se aprecia que el sexo determina la inserción por mecanismo informal (redes), siendo la proporción de ORA significativa para los hombres con un valor de 1,47 (IC95%= 1,08-1,98) sobre la referencia de mujeres.
- Se observa también que cuando el empleo no es una actividad laboral principal su asociación con la variable dependiente es preservadora. Esto es, una actividad laboral principal se obtiene con menor probabilidad mediante contactos personales. Otra lectura de este dato nos permite decir que si el empleo obtenido es una actividad laboral principal la probabilidad de encontrar trabajo mediante redes será de un 68% menor en comparación con una actividad laboral secundaria.

Figura 15. Visión gráfica de los coeficientes fijados.

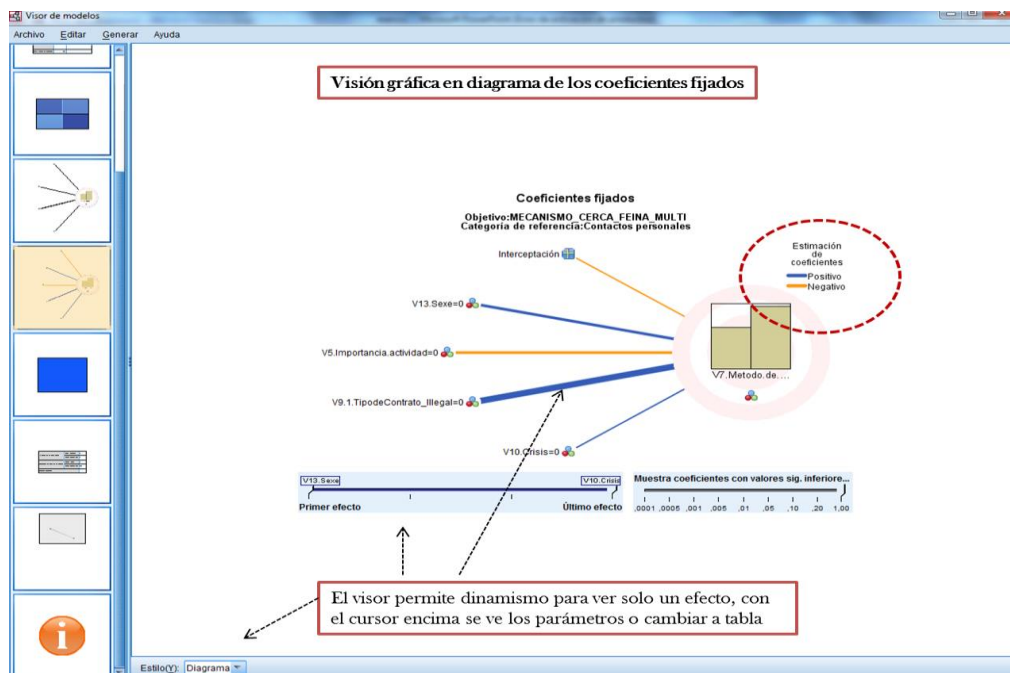
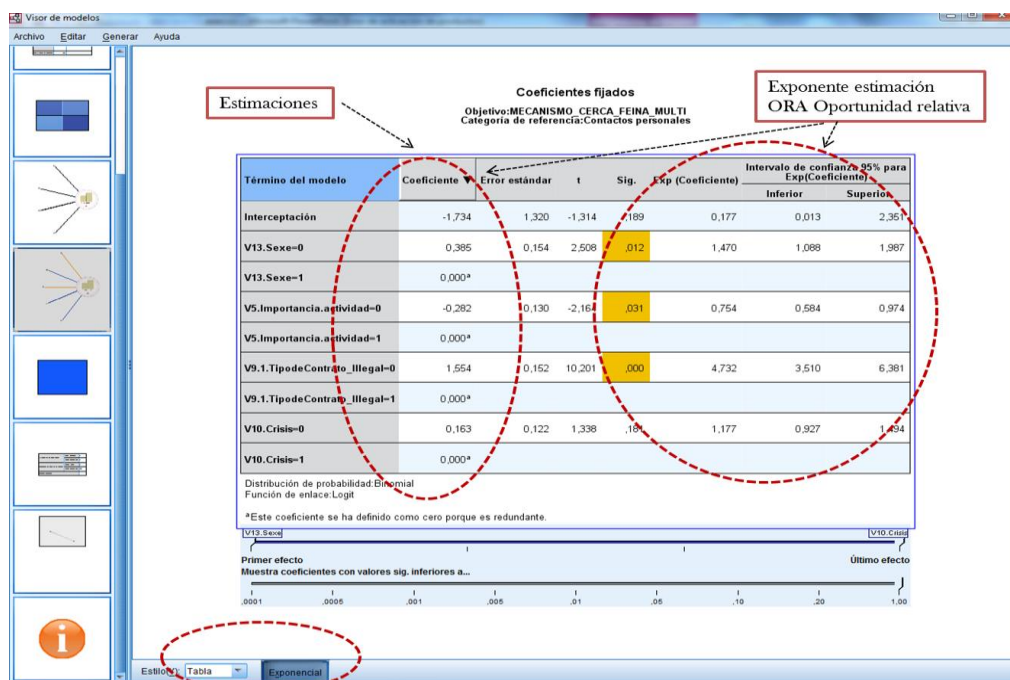


Figura 16. Distribución de las estimaciones y las ORA de los efectos fijados.



A través de la tabla de resultados (Figura 16) se pueden interpretar las estimaciones o las Oportunidades Relativas Asociadas (ORA), que son la elevación exponencial de las estimaciones. A nivel inicial tanto el sexo como la actividad principal o el tipo de contrato se asocian con la inserción por redes personales, siendo esta relación estadísticamente significativa:

- Se aprecia que el sexo determina la inserción por mecanismo informal (redes), siendo la proporción de ORA significativa para los hombres con un valor de 1,47 (IC95%= 1,08-1,98) sobre la referencia de mujeres.
- Se observa también que cuando el empleo no es una actividad laboral principal su asociación con la variable dependiente es preservadora. Esto es, una actividad laboral

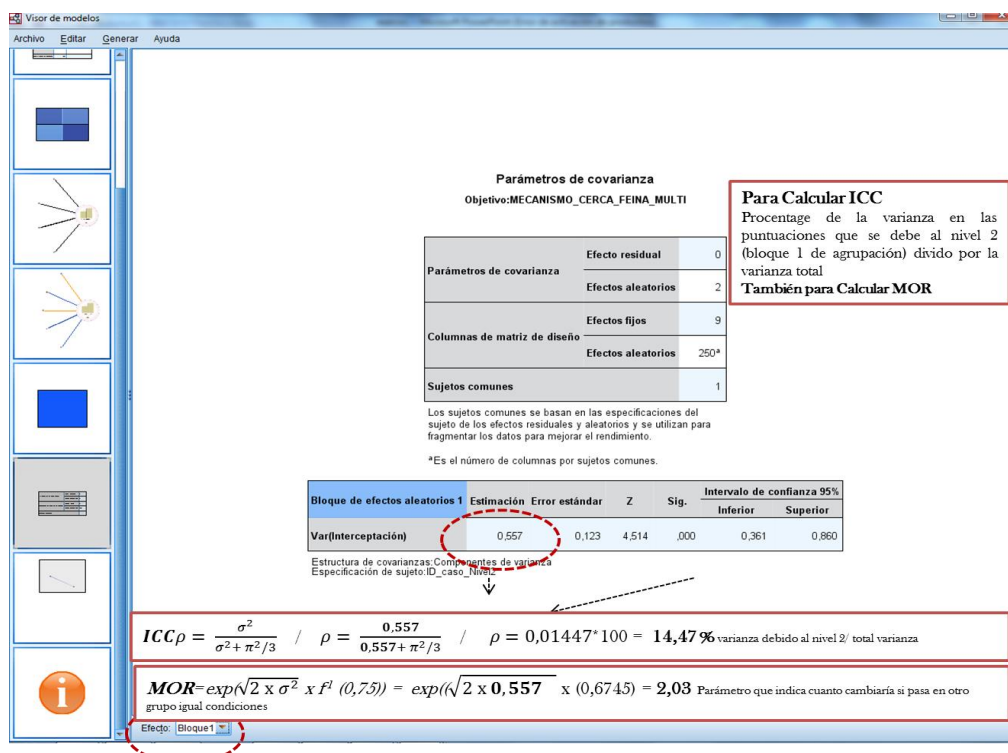
principal se obtiene con menor probabilidad mediante contactos personales. Otra lectura de este dato nos permite decir que si el empleo obtenido es una actividad laboral principal la probabilidad de encontrar trabajo mediante redes será de un 68% menor en comparación con una actividad laboral secundaria.

- La irregularidad del empleo es el factor explicativo más relevante para explicar la variable dependiente. Jóvenes que tienen una relación laboral sin contrato tienen una probabilidad 4,7 veces superior de haber obtenido el empleo mediante un contacto personal.

- Como hemos señalado precedentemente, el momento en que se ha obtenido el empleo (evento) de manera informal no es significativo, por tanto, la inserción por redes se manifiesta como una característica estructural del mercado del trabajo, pues no depende de la coyuntura económica.

Otra ventana interesante es la que facilita la información de los parámetros de covarianza y los efectos aleatorios. Gracias al valor del intercepto, si es significativo, se puede proceder a hacer la ecuación para calcular la ICC y la MOR (mediana de la ORA), que SPSS no ejecuta. Las ecuaciones se ofrecen encuadradas en la Figura 17.

Figura 17. Visor de los parámetros de covarianza de los efectos aleatorios.



El parámetro MOR es discutible en algunas disciplinas de las Ciencias Sociales, pero en epidemiología social y sociología de la salud se considera un buen indicador gráfico de la variabilidad entre grupos, ya que es una mediana que muestra cuanto varía la proporción de un grupo a otro si el individuo (o estructura del primer nivel) son similares. Es decir, ayuda a explicar la varianza del segundo nivel.

Como se aprecia en los resultados, al utilizar un enlace logit diferente a R, hay alguna pequeña diferencia en los resultados. En el caso de la ICC nos indica que la varianza debida al nivel segundo en relación a la varianza total es del 14,47%. Además, la proporción que ofrece la MOR es de 2,03.

3. Conclusiones

Se ha realizado una breve introducción a la lógica multinivel y se ha propuesto un ejemplo de aplicación de un modelo de regresión logística multinivel (RLM) a un estudio sobre la inserción laboral juvenil en el Área Metropolitana de Barcelona. Con este fin, se ha utilizado el procesador estadístico SPSS y el software R. Principalmente, el texto ayuda a comprender la lógica de base del análisis multinivel, observando las diferencias entre un procesador y otro en un modelo simple de dos niveles.

El procesador estadístico SPSS permite al investigador/a trabajar poco a poco el análisis multinivel a partir de su pregunta de investigación. Desde este punto de vista facilita el proceso secuencial desde el modelo nulo (modelo inicial sin variables predictores) hasta el modelo final y observar cómo van interfiriendo factores o covariables. Por el contrario, el entorno de trabajo R, por sus mismas características, propone un proceso de modelaje más rápido, centrándose únicamente en la definición de las variables independientes y la variable dicotómica dependiente.

En relación a la visualización de los resultados, SPSS ofrece al investigador/a centrar la atención en aquellos parámetros que desea observar, al mismo tiempo que facilita mayor dinamismo en la expresión gráfica y visual, si bien, el formato flash de visualización de los datos dificulta su posterior explotación en presentaciones u otras composiciones. En comparación, el entorno de trabajo R permite una visualización rápida de los resultados a través la consola, permitiendo una menor inversión de tiempo una vez adquirida la alfabetización básica del lenguaje y de la lógica multinivel.

4. Referencias

Andréu, J. (2011). El análisis multinivel: una revisión actualizada en el ámbito sociológico. *Metodología de Encuestas*, 13, 161–176.

- Murillo, F. J. (2008). Los modelos multinivel como herramienta para la investigación educativa. *Magis, Revista Internacional de Investigación en Educación*, 1, 45-62.
- Ohlsson, H., Beckman, A.; Johnell, K.; Hjerpe, P., Larsen, K. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health*, 60(4), 290–297.
- Bauer, D. J. y Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: Performance of alternatives specifications and methods of estimation. *Psychological Methods*.
- Berkhof, J. y Snijders, T. a. B. (2001). Variance Component Testing in Multilevel Models. *Journal of Educational and Behavioral Statistics*, 26(2), 133–152.
- Cebolla, H. (2013). *Introducción al análisis multinivel*. Madrid: Centro de Investigaciones Sociológicas (CIS).
- Heck, R. H. y Thomas, S. L. (2009). *An introduction to multilevel modeling techniques*. New York: Routledge.
- Heck, R. H, et al. (2012): *Multilevel modeling of categorical outcomes using IBM SPSS*, New York: Routledge.
- Pardo, A., Ángel, M., San, R. (2007). Cómo ajustar e interpretar modelos multinivel con SPSS. *Psicotherma*, 19(2), 308–321.
- Requena, F. (1991). *Redes sociales y mercado de trabajo. Elementos para una teoría del capital relacional*. Madrid: Centro de Investigaciones Sociológicas.
- Rieucan, G. (2008). Job advertisements and personal networks: two specific channels in the Spanish labour market. *Transfer: European Review of Labour and Research*, 14(3), 469-480.
- Smith, S. S. (2000). Mobilizing social resources: Race, ethnic, and gender differences in social capital and persisting wage inequalities. *The Sociological Quarterly*, 41(4), 509-537.
- Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canada Academy of Child and Adolescent Psychiatry*, 19(3), 227-229.